2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Identifying the Impact of Friends on their Peers Academic Performance

Philip Scanlon & Alan Smeaton
Insight Centre for Data Analytics
Dublin City University
Email: philip.scanlon@insight-centre.org

***Abstract*—Historically data collection in the research process involves either surveys, interviews or observation, or any combination of all three. Recent developments in the area of formative educational methods have enabled other data collection options. Data sources now available include university Virtual Learning Environments (VLEs), E-learning and many other knowledge management systems. Data-sets harvested from these sources are less susceptible to the inherent biases introduced through the intervention of human interpretation. Data is often structured, complete and traceable. The research in this paper aims to utilise one of these unique digital data-sets which represents the footprints created by student activities within a university environment and through Social Network Analysis to identify their influences within peer groups.**

I. Introduction

From the moment a student applies to a university, gains acceptance and attends the campus they are creating a unique digital footprint of themselves in the university IT systems.

A student's digital footprint can comprise of a number of components including:

- Demographic information
- Previous academic history
- Assignment and exam performance
- Library attendance and book withdrawls
- Access to WiFi or other networking systems

This research examines the digital footprint created by a student's use of the University WiFi system. The data being mined is the log files of requests by WiFi enabled device to access the wireless networking system, Eduroam.

In this work we will determine who becomes 'friends' with whom among a student population and thereby determine who is influencing whom in the learning process.

II. Background

The challenges in the examination of a micro-environment such as a university and the diverse elements that co-exist within that environment. Modern analysts have an ever-increasing number of approaches and numerous toolkits to carry out a core in-depth and complex study of this academic domain. One of the more recent evolutionary approaches is that of Learning Analytics.

A. Learning Analytics (LA)

As with many areas of analysis, learning analytics is based on historical data and is retrospective. Often a students' measure of success is based on the outcome at the end of a semester, and on the examination results alone which places a quantifiable measure on the previous semester's efforts.

Once an affiliated user enters a campus with an enabled device, it automatically requests connection to the Eduroam WiFi network. Each request is a unique identifier of who (device), where and when a device has requested access to through the wifi system. Figure 1 illustrates the number of student access requests to the Eduraom WiFi system for the month of March 2015 for our University.

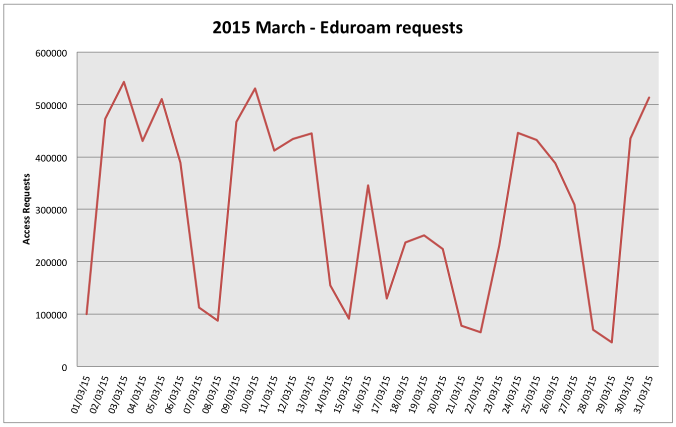

Fig. 1. Daily Eduroam Access requests.

III. Related Work

Rui Wang [1] carried out a *SmartGPA study* using students' direct reporting and passive sensing data collected from student smartphones, while endevouring to understand student behavioural patterns. He analysed the gps data, sensing data and audio data to understand the behaviour of students of academic different abilities. Through understanding different student behaviours he wished to determine if academic achievement can be predicted by student behaviour. Wangs' conclusion shows that there is correlation between GPA and sensed behaviours.

IV. Methodology

Our research relies on the premise that students on campus are engaged in either academic or social activities. Using a similar approach to that of Rui Wang [2] we sub-divided the campus into Academic and Social areas.

School rooms and Library are categoriesed as the *Academic* areas and cafes, shops, sports complex and Administration offices are classified as *Social*. As students pass between areas their device will seamlessly exit one NAS (access point device) range and be picked up by another. Their WiFi access request is captured, logged and stored. This data is the basis for our research.

A. Dataset Sample

The Dublin City University (DCU) student population is approximately 12,000 students per semester. Our research cohort is drawn from ten heterogeneous first year modules chosen from varying schools. In our test year these modules had 2,028 registered students at the commencement of the academic year and provide a representative cross section of students from various disciplines within the university.

For our research our goal was to identify co-located dyads, these are defined as a unique meeting between two WiFi enabled devices that are co-located during a specified time window. This time window is defined as being 20 minutes in duration. For example if two devices are connected to the same NAS (WiFi access point) within the same time window, they are deemed to be co-located and classified as a **meeting**. Our sample modules generated 22,800 pairwise dyads (meetings) in the first semester.

V. Findings

Having collected and consolidated our data we needed to extract useful knowledge. The primary question as outlined in Section I is based on the ability to identify *friends*. The premise is that *friends* are together in the same location, at the same time on a regular basis. It was therefore necessary to identify the degree to which students collocate and use this as a feature in our analysis.

We therefore extract useful features from the data. The features extracted included:

1) Degree of friendship (Pairwise), is a count of the number of unique friends each student meets during the semester.
2) Degree of interactions per location is the sum of all meetings per location, both academic and social, that a student has throughout the semester.

Using the extracted features and the semester-end exam results, we applied Linear Regression models, Pearson, and Ordinary Least Squares, plus machine learning techniques including Linear Regression and K- Neighbors Regression models to our data. This analysis was carried out on our sample modules, with very similar results across each.

A sample result from a single module is illustrated here. This module is taught in first year and in the first semester to a number of science degree programmes. A graphical representation in Figure 2 illustrates the findings of our regression analysis on this module.

We believe these results are indicative of the behaviour of new students in the first semester of the first year of their university career. Students will be adjusting to a new educational and social environments, taking part in new social

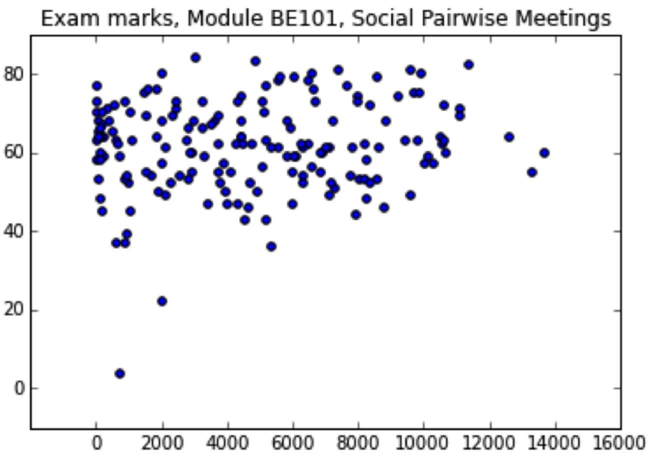

Fig. 2. Regression Analysis for BE101.

dynamics and in some cases will suffer from isolation from family and friends. For these reasons our initial results were not unexpected. Our network overview identified that students in their first semester do not tend to reach a stable number of ‘relationships’ until the 4th or 5th week of semester. Due to the group development process we believe that in the first semester of a programme, students have not established a group dynamic whereby they have any influence on their peers.

VI. Conclusion

We have utilised a non-invasive strategy using WiFi access log data to differentiate between the identity of a student and their peers. This research has determined that it is possible to identify students on a campus through the digital footprint provided by their WiFi activity. It has also concluded that, based on pairwise determination through co-location data, the degree of each student’s activity within the cohort of students can be determined. Through the use of Social Network Analysis techniques we have determined the degree of students within the selected cohort and carried out regression analysis with interesting findings.

VII. Future work

It is intended to carry out a longitudinal study utilising the initial students cohort to determine friendship patterns and the affect on academic results.

Acknowledgement

This paper is based on research conducted with the support of Science Foundation Ireland under grant SFI/12/RC/2289.

References

[1] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell. SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’15, pages 295–306, New York, NY, USA, 2015. ACM.

[2] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell. SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 295–306, New York, NY, USA, 2015. ACM.